IBM Applied Data Science Capstone Project by Coursera

Opening a new Coffee Café in Bengaluru, India

Girish A Koushik

July 2020

Introduction

“When you live in South India, coffee is your lifeline” [1]. This is how a regional newspaper article describes the coffee consumption in Bengaluru, which is the largest city in South India, the capital of the southern state of Karnataka and the ‘Silicon Valley’ of the Indian sub-continent. The article goes on quoting a number of people who would sometimes love to consume more than 5 cups of coffee a day, without worrying much about the medical research that presents the demerits of very high consumption of caffeine.

The coffee shops in Bengaluru are frequented by customers who cherish either the ‘traditional filter coffee’ or the cappuccino, espresso and other ‘international variants’ of coffee. The most recent trend in the city’s coffee culture stems from the growing emphasis on artisan coffee and the perception of coffee as an experience rather than just a daily dose of caffeine. This comes from an increasing interest in quality coffee among the millennial who look at coffee consumption as an aspirational experience, a new lifestyle, especially as far as out-of-home coffee consumption [2].

The southern region of India accounts for the largest share of coffee in the beverage market with 73 percent. In 2020, the number of organized café chains across India was forecast to increase to around 6,200, up from about 3,500 in 2017. The value of coffee in the retail market across India was forecast to be 140 Billion Indian Rupees in 2020 [3]. To cater to the demand of the working population, new coffee shops are being opened across the city and hence, the location of the coffee café is one of the most important decisions that will determine whether the café will be a success or not.

Business Problem

The objective of this capstone project is to analyze and select the best locations in the city of Bengaluru, India to open a new coffee café. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question: In the city of Bengaluru, if a business owner is looking to open a new coffee café, where would you recommend that they open it?

Target Audience of the Project

The findings of this project are not only useful to small business owners who are looking to open a new coffee café and/or expand their business across their city but also to popular coffee chains such as Starbucks, Barista, Café Coffee Day etc., who are looking to increase their customer base.

Data Requirements

To solve the problem, we require the following data:

1. List of neighborhoods in Bengaluru. This defines the scope of the project, which is confined to the city of Bengaluru (capital of the state of Karnataka) in India.
2. Latitude and Longitude of the above neighborhoods. This is required in order to obtain the venue data and to plot the map.
3. Venue data, particularly related to the coffee shops. We will use this data to perform the clustering on the neighborhoods.

Data Collection

The Wikipedia page (Category: Neighborhoods in Bangalore) contains the list of neighborhoods in Bengaluru having a total of 128 neighborhoods. Web scraping techniques will be used to extract the data from the Wikipedia page, with the help of Python requests and 'beautiful soup' packages. The python 'Geocoder' package will retrieve the latitude and longitude coordinates of the neighbourhoods.

Further, the Foursquare API will be used to get the venue data for the above neighbourhoods. Foursquare API will provide many categories of the venue data and the project interest will be the 'Coffee Shops' category in order to solve the business problem. This project that will make use of many data science skills, from web scraping (Wikipedia), working with API (Foursquare), data cleaning, data wrangling, to machine learning (k–means clustering) and map visualization (Folium).

Methodology

First, the list of neighborhoods in the city of Bengaluru is obtained from the Wikipedia page which can be found here. The web scraping will be done using Python requests and 'beautiful soup' packages to extract the list of neighborhoods. However, this is just a list of names of the neighborhoods. The geographical coordinates of the neighborhoods are obtained in the form of latitude and longitude. To do so, the python 'Geocoder' package is used that will allow the conversion of address. After gathering the data, the data is populated into a pandas data frame and then visualize the neighbourhoods in a map using the Folium package. The data frame containing the latitudes and longitudes are merged with the data frame containing the names of the neighborhoods.

Next, the Foursquare API will be used to get the top 100 venues that are within a radius of 2 kms. In order to obtain the Foursquare Client ID and Foursquare Client Secret Key, a Foursquare Developer account need to be registered. API calls are made to Foursquare by passing the Client ID, Secret Key and API version, to obtain the geographical coordinates of the neighborhoods in a python 'for' loop. Foursquare will return the venue data in JSON format and using that one can extract the venue name, venue category, venue latitude and longitude. With the above data, the number of venues returned can be checked for each neighbourhood and can examine how many unique categories can be curated from all the returned venues. Then, each neighbourhood will be analyzed by grouping the rows by neighbourhood and taking the mean of the frequency of occurrence of each venue category. By doing so, the data is prepared for use in clustering. Since, the project is based on coffee shop / coffee café, the 'Coffee Shops' data as venue category is filtered for the neighbourhoods.

Lastly, clustering is performed on the data by using k-means clustering. k–means clustering algorithm identifies 'k' number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible. It is one of the simplest and most popular unsupervised machine learning algorithms and is particularly suited to solve the problem for this project. The neighborhoods will be divided into 3 clusters based on their frequency of occurrence for 'Coffee Shops'. The results allow the user to identify the neighborhoods which have higher concentration of coffee shops and the neighborhoods which have fewer number of shopping malls. Based on the occurrence of coffee shops in different neighborhoods, the question as to 'which neighborhoods are most suitable to open new coffee shop' can be answered.

Results

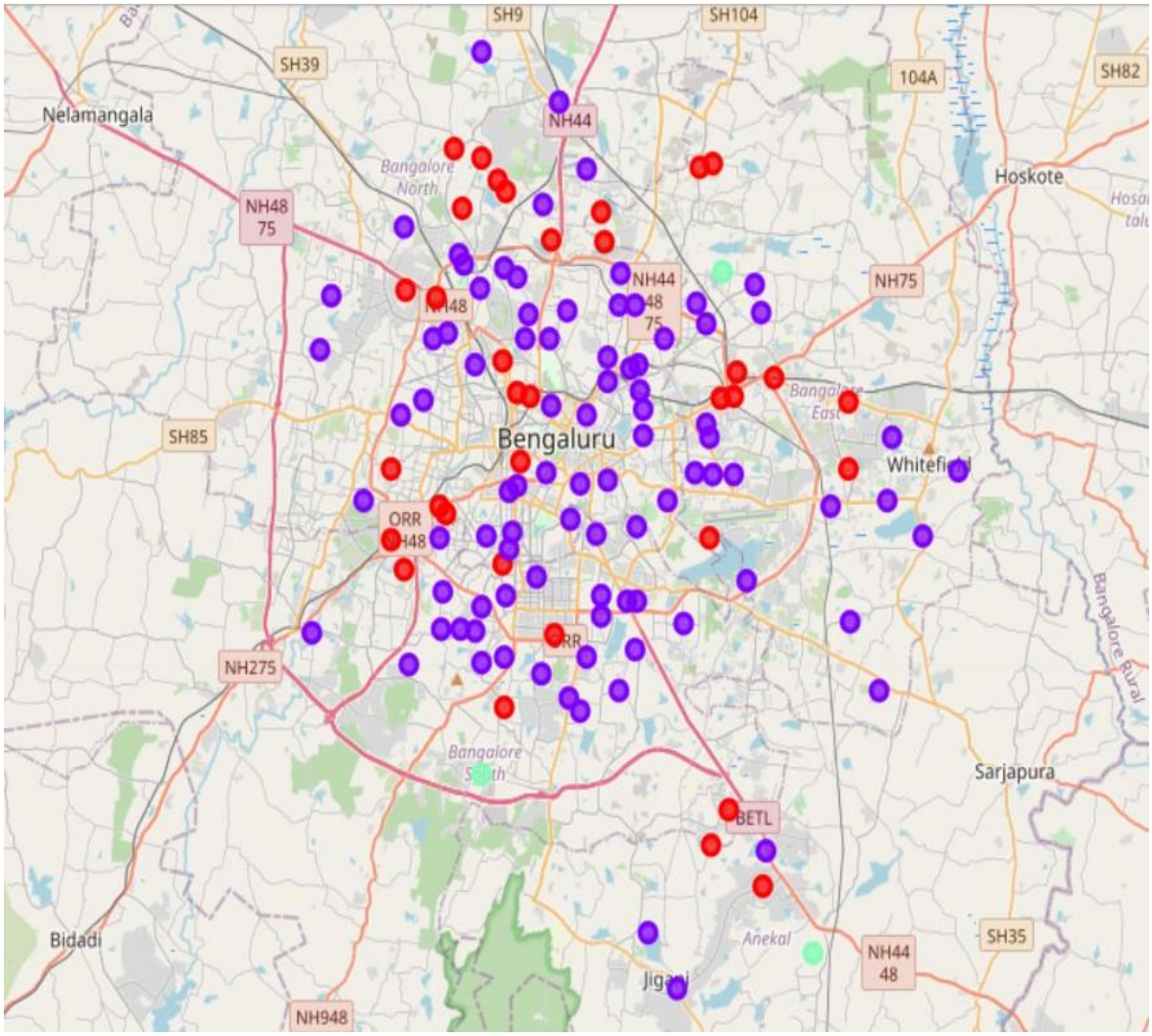
The k–means clustering results show that the neighborhoods in Bengaluru can be divided into 3 clusters based on the frequency of occurrence for ‘Coffee Shops’.

Cluster 0: Neighborhoods with moderate number of coffee shops

Cluster 1: Neighborhoods with very low number to no coffee shops

Cluster 2: Neighborhoods with high number of coffee shops

The results from the clustering are visualized in the map with cluster 0 in ‘red’ color, cluster 1 in ‘purple’ color and cluster 2 in ‘mint green’ color.



Observations and Recommendations

Most of the coffee shops are concentrated in the southern, eastern and central parts of the Bengaluru city, with the highest numbers in cluster 2 and moderate numbers in cluster 0. On the other hand, cluster 1 has very low number to no coffee shops in the neighborhoods. This cluster represents a great opportunity and high potential areas to open new coffee shops as there is very little to no competition from the ones already present. Meanwhile, coffee shops in cluster 2 are likely suffering from intense competition due to high concentration of other coffee shops. Therefore, this project recommends coffee business owners and popular coffee chains to capitalize on these findings to open new coffee shops in the neighborhoods in cluster 1. Coffee business owners with unique selling point (USP), stand out from the competition and can also open new coffee shops in the neighborhoods in clusters 0 with moderate competition.

Conclusion

In this project, the business problem has been identified, specifying the data required, collecting and preparing the data, performing machine learning by clustering the data into 3 clusters based on their similarities, and finally providing recommendations to the relevant stakeholders i.e., coffee business owners and popular coffee chains regarding the best locations to open a new coffee café. To answer the business question that was raised in the business problem section, the solution proposed by this project is: The neighbourhoods in cluster 1 with low frequency of occurrence are the most preferred locations to open a new coffee café. The findings of this project will help the relevant stakeholders to capitalize on the opportunities on high potential locations while avoiding neighborhoods with high concentration of coffee shops in their decision to open a new coffee café.

Future Research

In this project, only one feature was considered i.e., the frequency of occurrence of coffee shops. There are other factors such as population and locations where new corporate offices are coming up. As a special case with international coffee chains (Starbucks, Barista etc.), people with different income groups are useful as these coffee chains tend to be more expensive. Using these data could significantly improve the accuracy of the prediction model. It is also possible that many small coffee shops which are already present are not available in the

Foursquare API. However, such data were not available while carrying out this project and future research could devise a strategy to incorporate such data to be used in the clustering algorithm to determine the preferred locations.

References

[1] https://bangaloremirror.indiatimes.com/bangalore/others/how-many-cups-of-coffee-do-you-drink-do-tell-with-hand-on-heart/articleshow/69314831.cms

[2] https://www.deccanherald.com/metrolife/bengaluru-s-coffee-culture-has-changed-tremendously-746249.html

[3] https://www.statista.com/topics/4590/coffee-market-in-india/

[4] Neighborhoods in Bangalore → Wikipedia (https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Bangalore)

[5] Beautiful Soup Documentation → Beautiful Soup (http://beautiful-soup-4.readthedocs.io/en/latest/)

[6] Pandas Documentation → Pandas (https://pandas.pydata.org/docs/)

[7] Foursquare Developer Documentation → Foursquare (https://developer.foursquare.com/docs)

Appendix

The highlighted areas depict the possible locations for opening a coffee café.

Cluster 0: Neighborhoods	Cluster 1: Neighborhoods	Cluster 2: Neighborhoods
Yeswanthpur	Nandini Layout	Chandapura
Shikaripalya	Nagarbhavi	Horamavu
Domlur	Muthyalanagar	Anjanapura
Seshadripuram	Murphy Town, Bangalore	
Electronic City	Milk Colony	
Kundalahalli	Suddaguntepalya	
Ramachandrapura	Marathahalli	
Rajarajeshwari Nagar, Bangalore	Padmanabhanagar	
Hebbal	Whitefield, Bangalore	
Hoodi	Mahalakshmi Layout	
Peenya	Wilson Garden	
Palace Guttahalli	Yelachenahalli	
J. P. Nagar	Madiwala	
Nayandahalli	Lingarajapuram	
Mariyannapalya	Mathikere	
Kalyan Nagar	Vyalikaval	
Malleswaram	R. T. Nagar	
Mahadevapura, Bangalore	Puttenahalli	
Kodigehalli	Tannery Road	
Konanakunte	Statue of Queen Victoria, Bangalore	
Dasarahalli	Statue of Edward VII, Bangalore	
Chickpet	Siddapura, Bangalore	
Krishnarajapuram	Taverekere	
Vidyaranyapura	Shivajinagar, Bangalore	
Vijayanagar, Bangalore	Sanjaynagar	
Ananthnagar	Sadashivanagar	
Bagalur, Bangalore Urban	Richmond Town	
Byatarayanapura	Ramanjaneyanagar	
Brookefield	Ramamurthy Nagar	
Bommasandra	Ramagondanahalli	
Thindlu	Ulsoor	
Thyagarajanagar	Rajajinagar	

Bharathnagar	Uttarahalli	
Vimanapura	Varthur	
	Sulikunte	
	Vasanth Nagar	
	Sahakara Nagar	
	Adugodi	
	Koramangala	
	Gandhi Bazaar	
	Fraser Town, Bangalore	
	Ejipura	
	Dollars Colony	
	Devarachikkanahalli	
	Cox Town, Bangalore	
	Cooke Town	
	Chikkalasandra	
	Chamrajpet	
	Carmelaram	
	CV Raman Nagar	
	Bommanahalli	
	Bilekahalli	
	Bellandur	
	Basaveshwaranagar	
	Basavanagudi	
	Banaswadi	
	Banashankari	
	Bahubalinagar	
	Babusapalya	
	BTM Layout	
	Austin Town	
	Arekere	
	Kumaraswamy Layout	
	Ganganagar, Bangalore	
	Gandhi Nagar, Bangalore	
	Gowdanapalya	
	Koppa gate	
	Kodihalli, Bangalore	
	Kengeri	
	Yelahanka	
	Kamakshipalya	
	Kalkere	
	Kalikanagar	
	Kalasipalyam	
	Kaggadasapura	

	Jigani	
	Jeevanbheema Nagar	
	Girinagar	
	Jayanagar, Bangalore	
	Jakkur	
	Ittamadu	
	Indiranagar	
	Immadihalli	
	Hulimavu	
	Hebbagodi	
	Hanumanthanagar, Bengaluru	
	HSR Layout	
	HBR Layout	
	Jalahalli	
	Kammanahalli	